汉语儿童读物的质性分级标准

李燕[1]，郭梓楠[1]，吴昊旻[2]，郭文[2]，刘嘉[2]

（1. 北京师范大学心理学部，北京 100875；

2. 清华大学脑与智能实验室，北京 100084）[1]

摘要：小学生处于学会阅读的关键期，其语言和认知成长需要大量难度适宜、利于理解的阅读材料。当前的阅读分级主要采用字、词、句等浅表层面的量化文本指标，较难适应汉语表意性等特点和儿童认知发展需求。本研究首先采用元分析等方法，归纳和筛选出文体、主题、角色、情节、语言特点、文本结构、文化知识以及生活经验 8 个质性分级指标，然后在国际上首次采用 SOLO 分类评价法来为每个指标制定契合儿童认知发展结构的等级水平描述，最终形成《汉语儿童读物的质性分级标准》及操作手册。进一步的实证实验结果说明，该分级标准具有较高的区分效果、评分者一致性信度和预测效度，能够客观稳定和科学可靠地评估儿童读物的文本复杂性。综上，此标准可有助于教师、出版编辑和家长评估儿童读物的适用年级，利用汉语分级阅读来促进儿童语言与认知发展，实现独立阅读。

关键词：分级阅读；儿童读物；文本复杂性；质性指标；SOLO 分类评价法

chinaXiv:202301.00068v1

基金项目：国家自然科学基金委国际（地区）合作与交流项目“成为面孔识别专家需要什么样的经验：基于逆向工程方法的研究”（31861143039）。

收稿日期：2023 年 1 月 3 日

作者简介：李燕，北京师范大学心理学部博士，主要从事教育和心理测量、分级阅读研究；郭梓楠，北京师范大学心理学部硕士研究生，主要从事心理测量和汉语分级阅读研究；吴昊旻，清华大学脑与智能实验室硕士研究生，主要从事文本分析与阅读分级研究；郭文，清华大学脑与智能实验室科研助理，硕士，主要从事学习分析与阅读分级研究；刘嘉，清华大学基础科学讲席教授、清华大学心理系系主任，主要从事人工智能的认知神经基础、视觉智能研究，通讯作者 Email: liujiaTHU@tsinghua.edu.cn。

第一作者联系方式：李燕 13220027976（微信同号） liyanug@126.com

对处于阅读关键期的小学生来说，分级阅读被公认为是一种因材施教、卓有成效、值得推广的阅读教育体系[1]。大量研究揭示，难度适宜的阅读材料能作为儿童阅读能力发展的关键脚手架，利于培养儿童的阅读技能、策略、专注力和习惯[2]。因此，我国统编版小学语文教科书和教育部基础教育课程教材发展中心发布的《中小学生阅读指导目录(2020 年版)》等均体现了阅读分级理念，倡导应根据不同学生的情况，适时推荐文化品位较高、难易程度适当的课外读物[3-4]。

调查显示，尽管我国少年儿童的阅读需求强劲，每年新出版的少年儿童读物超 2 万余种，但是大量童书存在"缺少适读推荐、适读范围过宽、适读类型僵化、推荐年龄错误"等问题。这是因为我国阅读教育中缺乏一套科学、全面、本土化的阅读分级标准，仅有出版机构和研究机构提出的一些初步规范[5-6]，因此较难发挥出对儿童阅读的有效指导和教育价值。因此，亟需建立符合中国儿童特点的读物分级标准，以开展和推广我国儿童读物的分级，助力引导学生的阅读发展。

一、阅读分级

（一）当前阅读分级的现状与不足

当前，阅读分级研究集中在对文本可读性的量化分析领域，其原理为借助自然语言处理（natural language processing，NLP）技术来自动测量单词、句子、结构等层面的文本指标，并通过公式模型来估计文本的可读性（readability)[7]。如著名的蓝思公式将词频、句长指标作为文本难度的衡量标准。我国研究者也发现，汉字种、词种难度和虚词比例可作为小学汉语课文册数的最佳预测指标[8]。此类可读性模型实施便利，能较好地估计因字词难度产生的阅读认知负荷，已在教材编写、童书编辑和出版等领域得到了较为广泛的应用[9]。

然而，传统的量化分析法几乎完全依赖于可被自动提取的浅层、微观语言特征，仅能解决"可读性"问题，难以应对儿童读物的"复杂性"问题。多种语言的研究发现，语言作品的阅读难度不仅在于字、词、句等浅表层面的解码（decoding），更在于其所蕴含的情节发展、思想深度、文化底蕴等宏观层面的理解（comprehension）[10]。相较于从字、词、句中获取语义信息，研究者普遍认为文本理解需要读者进行推理、类比、审辩、反思等复杂加工[11]，这种积极的意义建构过程对儿童认知和语言发展至关重要。同时，文本复杂性也并不绝对，而是相对于儿童自身的背景知识和生活经验来变化的，会因语言、认知、心理发展水平和阅读任务而产生不同的理解层次。因此，在对文本可读性进行量化分析的基础上，还需要针对儿童读物的复杂性进行研究，以形成全面的和系统的分级指标体系。

此外，由于汉语的特殊性，针对浅表层面解码的量化指标对汉语阅读分级的适用性较差。这是因为汉语是一种高度完备的表意文字，其表意性体现在汉字的形旁部件能规则化地传达语义信息，形成"触类旁通"、"举一反三"的阅读效果，进而降低了字词难度对汉语阅读理解的影响[12]。同时，作为基本意义单元的汉字能较为灵活地构成词语，因而仅 3500 常用汉字就可以覆盖日常生活中 99%以上的文字[13]，其效价远高于英语词汇（99%覆盖率需认识 10000 词左右）[14-15]。因此有研究发现，仅通过字词难度、虚词比例等量化指标来预测汉语文本可读性对中高年级文本难度预测准确率不高，同时普适性（generality）较差，即量化文本指标难以在新的语料中保持原有的预测力[8]。

chinaXiv:202301.00068v1

因此，汉语儿童读物阅读分级的指标体系须在量化文本指标的基础上，充分契合汉语特有的语言特点和学习规律。综上所述，我们提出基于文本复杂性的质性分级是构建完善的汉语阅读分级的关键。

（二）汉语儿童读物的质性分级解决方案

研究显示，质性分级是儿童读物复杂性问题的有效解决方案。质性分级的原理是由审慎的读者按照一定的标准或规则，对图书的文本特征和适用对象进行评级的分析方法[16]。这种由教师、编辑及家长们对书籍难度进行质性估量的方法具有广泛的应用，如美国共同核心州立标准（CCSS）中的文本复杂度评估[17]、新英格兰共同评估（NECAP）[18]、美国大学入学考试（ACT）[19]、中国中小学英语分级阅读项目[20]以及上海市中小学汉语分级阅读标准研制项目[6]等，都将质性评估纳入了分级阅读教育体系中。

和量化分析常使用的字、词、句等文本指标不同，质性分级关注书籍的主题内容、思想深度、文本结构、背景知识需求等更为深层的语言特征。这些语言特征目前难以通过 NLP 自动提取，需通过标准化的质性分析范式来主观评估。例如，CCSS 采用的量规与样例（rubrics and exemplars）法将这些指标放置在一个连续量尺上，量尺的一端是特征的初级水平，另一端是复杂和高级的水平，而量尺的“刻度”则是相应指标水平的描述和文本样例[16]。此类质性评级方法实现了深层次语言特征的判定，其标准化范式提升了质性方法的客观性，为读者的主观判断提供了锚定依据。

但就构建汉语阅读分级的指标体系而言，上述研究存在一定的不足。在研究对象方面，上述研究大都基于英语材料或汉语课文语料，缺少基于汉语儿童读物的研究，指标体系的本土化和适用性欠佳。在研究方法上，相关研究的指标选取和等级水平的界定仅基于评审者的个人主观经验，存在指标等级模糊、不易评定、主观性较高的问题。在效果验证上，这些质性评估往往缺乏信效度检验，进而缺少可信度和可重复性等证据[21]。

此外，如何让阅读分级反映儿童的语言、思维发展阶段和需求也是一个不容忽视的问题。阅读分级的本质是为了推荐适合儿童最近发展区的图书以促进儿童的独立阅读发展，则汉语阅读分级的关键科学问题不只是选取和评估文本特征，更在于如何基于儿童语言和认知发展情况来预测其阅读各类书籍的学习效果。著名儿童心理学家皮亚杰的发生认识论指出，儿童的认识发展具有阶段性，不同阶段的认知水平有质的不同，对某种特定内容或行为的学习，只有当存在相应的结构时才能发生[22]。而比格斯等人在此基础上提出了 SOLO（structure of observed learning outcomes）分类评价法，进一步将儿童对具体学科知识、学习活动的认知过程概括为发展阶段和思维结构，即前结构、单点结构、多点结构、关联结构以及抽象拓展结构[23]。大量研究证实，SOLO 分类的 5 层次思维结构可以有效解释学生在语文、数学、科学等多学科任务的反应水平，进而描述学生认知发展的普遍规律[24]，可以作为儿童读物分级的理论基础。因此，我们采用 SOLO 分类法构建质性文本指标的等级水平描述，以自上而下地从儿童认知结构和理解能力的角度描述文本指标的层次水平，进而解释儿童的阅读学习效果和内因，增强质性分级的科学性和教育价值。

因此，本研究广泛调查了国内外教育学、心理学、儿童读物出版等领域的质性分级研究和实践

chinaXiv:202301.00068v1

经验，通过元分析归纳总结了影响汉语文本复杂性的质性指标，并采用 SOLO 分类法形成了各项指标的标准化等级水平，最终研制了一套《汉语儿童读物分级的质性评估标准》（以下简称《分级标准》）。此外，我们还采用心理测量效度框架开展了双评实验，系统验证了《分级标准》的区分效果、评分者一致性信度和预测效度。

二、研究方法

（一）研究对象

本研究选取了 150 本质量优秀的儿童书籍，来自教育部发布的《中小学生阅读指导目录（2020 年版）》或国内教育名家、名校推荐两次以上的小学分级阅读书目。书目中不仅包含中外各类优秀文学作品，还加入一定比例的图画书和实用作品，如介绍自然科学、人文社会、哲学、艺术等学科的普及性读物。这些汉语儿童读物都具有专家论证的适宜阅读年级，因此我们将适宜阅读年级的均值作为读物的文本复杂性等级（取整为 1-6 级），作为儿童读物质性分级的金标准。

（二）质性分级标准的构建

分级标准的构建分为三个部分，指标初筛、初评实验及形成标准（图 1）。

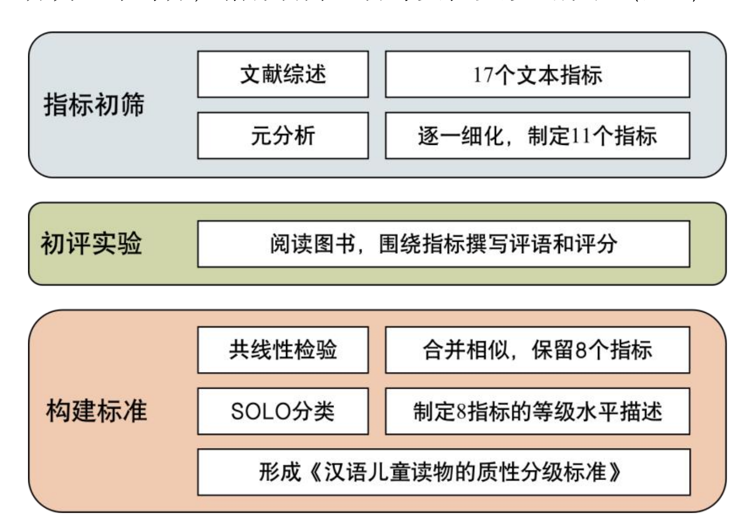


图 1 质性分级标准的构建流程

指标初筛。通过分析教育学、心理学和童书出版领域的相关文献，我们对国内外的主流质性指标体系进行归纳，初步筛选出 17 个文本指标。采用基于团队讨论的元分析法逐一细化了这些指标的定义和内涵，制定了容易理解、切实可评、与文本复杂性关联较大的 11 个指标。

初评实验。为了形成符合汉语儿童读物特点的指标等级，我们招募 13 名审慎的读者阅读了这些儿童读物。这些读者需分别围绕这 11 项指标来对这些图书撰写评语，并使用文中的内容和细节作为支撑。同时，我们还鼓励读者按照自己的感受对图书的指标水平打分。

形成标准。在初评实验结果的基础上，我们通过相关分析检验出了存在共线性问题（相关>0.80）的质性指标。为了减少这些指标含义的重叠，我们将 11 个指标梳理合并为 8 项指标，即文体、主题、角色、情节、语言特点、文本结构、文化知识、生活经验。除了汇总了自下而上的儿童读物指标评

chinaXiv:202301.00068v1

语，我们还参考 SOLO 分类法中的学生思维结构来自上而下地归纳学生阅读各级读物时需要的认知水平特点，以作为这些指标的等级水平描述。综上，研究团队撰写了《汉语儿童读物的质性分级标准》和《汉语儿童读物的质性分级操作手册》。

（三）质性分级标准的实验验证

为进一步验证《分级标准》的客观性和有效性，本研究设计了儿童读物的双评验证实验。具体来说，本研究招募了 22 名教育、心理、中文、文学专业的语文教师、出版从业者、在校研究生和高年级本科生作为评分员。22 名评分员被随机分为 11 组，每组阅读并评价 18 本相同的儿童读物，最终形成 191 对评分结果（其中 1 组评分员仅完成 11 本书的双评，每本书被评价 2 – 4 次）。评分员在完整阅读儿童读物后，依据研究者提供的材料《分级标准》和《操作手册》来对 8 个文本指标评分（1 – 5 级），并撰写具体的评分理由。

为了保证评级质量，我们使用了多种设置来控制实验的干扰因素。实验前，研究者对全部评分员进行了集体培训，包括解读打分标准及指标内容、讲解读物评分样例、现场练习及答疑等，确保评分员充分理解评分规范及流程。其次，评分员需完整阅读儿童读物后才可以评分，且需要结合读物的细节来撰写打分理由，这可以保证评分结果是客观真实的。最后，评分员们全程独立完成实验，既不知道读物的阅读难度等级，也不清楚其他人的待评书目，这可以保证评分结果的独立性。

（四）数据分析

本研究根据 11 组评分人对分级标准中难度等级和 8 个文本指标评分结果，使用 SPSS 和 Python 软件进行统计分析，分别计算了每组评分者的皮尔逊积差相关系数和斯皮尔曼等级相关系数作为评分者间一致性信度指标，计算了 Cronbach alpha 系数作为评分者内一致性信度指标。此外，本研究使用多元回归来估计了本阅读分级标准的预测效度。

三、结果

（一）读物复杂性的质性分级指标体系

为了建立一个全面、有效的文本复杂性质性评估指标体系，本研究首先对国内外的代表性研究[5-6,17-20]进行检索和分析，归纳了用语不同但内涵相似的指标（如主题、内容等），并剔除了与文本复杂性关联不紧密或不易评价的指标（如文本长度、写作手法、插图特点、排版形式等）。如表 1 所示，我们初步形成了 11 个质性评估指标，篇章类指标包含主题内容、人物角色、情节、文本类型/体裁；语言类指标包含词汇、句子结构、文本结构、语言常规性和清晰度；背景知识类指标包含文化知识、互文性、生活经验。

我们采用相关分析检验了这些指标在初评实验数据中的共线性，合并了相关系数 > 0.80 的文本指标，即将词汇、语言常规性和清晰度、句子结构概括为语言特点，将文化知识和互文性概括为文化知识。由此，本《分级标准》的文本复杂性指标共包含 3 个方面的 8 个指标。这些指标由浅入深地反映了儿童读物的浅表语言特征、深层思想特征和对儿童群体的认知、心理发展需求，因此作为汉语儿童读物分级的主要维度。

chinaXiv:202301.00068v1

chinaXiv:202301.00068v1

表 1 汉语儿童读物分级的质性评估指标

一级指标	参考指标及依据	最终指标
篇章	文本类型/体裁[6,17,20]	文体
	主题内容[5,6,17-20]	主题
	人物角色[6,20]	角色
	情节[6,20]	情节
语言	词汇[6,17,19]	语言特点
	句子结构[17,20]	
	语言常规性和清晰度[17-18]	
	文本结构[6,17-20]	文本结构
背景知识需求	文化知识[17-18]	文化知识
	互文性[6,20]	
	生活经验[6,17]	生活经验

（二）汉语儿童读物的质性分级标准

构建好文本指标后，我们根据 SOLO 分类法和初评中的儿童读物指标评语，为每个指标编制了 4－5 个水平的等级描述（1 级最易，5 级最难）。例如，根据 SOLO 的 5 水平思维结构（前结构、单点结构、多点结构、关联结构以及抽象拓展结构），情节指标的 5 个水平依次为：1 单一情节，2 重复情节，3 连续情节，4 多情节，5 复杂、多线索情节。由此最终形成《汉语儿童读物的质性分级标准》（表 2）。

为了保证读物分级的质量，我们还编制了《汉语儿童读物的质性分级操作手册》，涵盖了更为详细的指标介绍、评级原则、注意事项、相似等级的区分方法、常见问题以及评级样例等内容。例如，我们发现评分员可能会混淆情节指标中的水平 2（重复情节）和水平 3（连续情节），特在《操作手册》中界定了这两个水平的异同："重复情节与连续情节都具有多个情节。前者的连续性不强，并且情节之间相对独立或有一定重复。后者的情节相互衔接、递进，推动故事的发展和深入"。完整操作手册可联系研究团队获取。

表 2 汉语儿童读物的质性分级标准

指标	水平 1 前结构	水平 2 单点结构	水平 3 多点结构	水平 4 关联结构	水平 5 抽象拓展结构	评分
文体	1. 图画书，童谣，儿歌。	2. 童话，寓言，含简单的标牌、图示等的说明性文字。	3. 叙事性文本，如故事，短篇小说；图文结合的实用文，简单说明	4. 文学作品，如历史故事、戏剧、散文、诗歌、中长篇小说；纪实文	5. 专业类、科技类说明文，议论文，传记等。	

chinaXiv:202301.00068v1

			文、科普文。	章。		
主题	1. 简单日常主题：日常生活中的主题及概念（如家庭、动物、日用品、交通工具、食物、身体、颜色、游戏等）。	2. 有教育意义的日常主题：日常生活相关，且有利于积极心态、习惯和品质形成的主题（如快乐、分享、合作、尊重等），以及人际交往主题（如亲友、师生、陌生人等）。	3. 日常主题+少数非日常主题：既有与日常生活相关的主题，也有少数超出日常生活的主题（如冒险故事）。	4. 非日常主题：有超出个人经历和日常生活范围的主题出现，需要一定推理能力才能理解的抽象主题（如奇幻、探险、魔法、侦探等）。	5. 社会科学类非日常主题：社会类（如社会争议、人文艺术、道德伦理等），科学类（如科普百科、科技进展）以及健康类（如家庭问题、衰老与死亡、身心障碍等）。	
角色	1. 简单	2. 较简单	3. 较复杂	4. 复杂多面	5. 复杂变化、象征化	
情节	1. 单一情节	2. 重复情节	3. 连续情节	4. 多情节	5. 复杂、多线索情节	
语言特点	1. 白话，词汇简单且口语化，大部分为简单句，韵律感强，语言直白易理解。	2. 白话，词汇简单且口语化，以简单口语叙述为主，大部分为简单句，包含较短分句。有少量不影响理解的夸张、比喻、拟人等修辞手法，易理解。	3. 白话，词汇有一定难度且书面化，偶有长句、复杂的句子，以简单书面语叙述为主，含夸张、比喻、拟人等修辞手法，易理解。	4. 大部分是白话，以描述性书面语为主，夹杂一些较为拗口、晦涩的语言（如文言、方言），句子结构复杂、多变，不影响全文意义的理解。	5. 大部分是白话，书面化难词、长句较多，句子结构复杂、多变，多处出现较复杂的抽象、比喻（暗喻）、讽刺等手法，使全文意义深刻。	
文本结构	1. 结构清晰，叙述具有重复性，很容易预测。	2. 文本结构清晰，按时间或事物发展顺序叙述，较容易预测。	3. 包含一些复杂结构，按时间或事物发展顺序叙述，较容易预测。	4. 叙述方式多样，叙述过程中发生时间转换，偶尔出现较难预测的故事走向。	5. 叙述方式多样，叙述过程中发生时间转换，难以预测故事走向。	
文化知识	1. 包含日常使用知识，不涉及任何历史、文化、学科专业知识。	2. 所涉及的历史、文化、学科知识，为日常生活中可了解的常识。	3. 所涉及的历史、文化、学科知识有一定专业性，可通过阅读上下文理解。	4. 所涉及的历史、文化、学科知识有一定专业性，需要积累一定阅读经验才能理解。	5. 所涉及的历史、文化、学科专业知识深奥，脱离日常生活，有较大理解难度，需查找资料来辅助理解。	
生活经	1.文中描述的是日常生活经历或普遍社会经	2.文中描述与一般人日常生活经历相似，少量涉及地域、文化、	3.文中描述与一般人日常生活经历联系较少，有特殊	4.文中描述与一般人日常生活经历脱离，有独特的人		

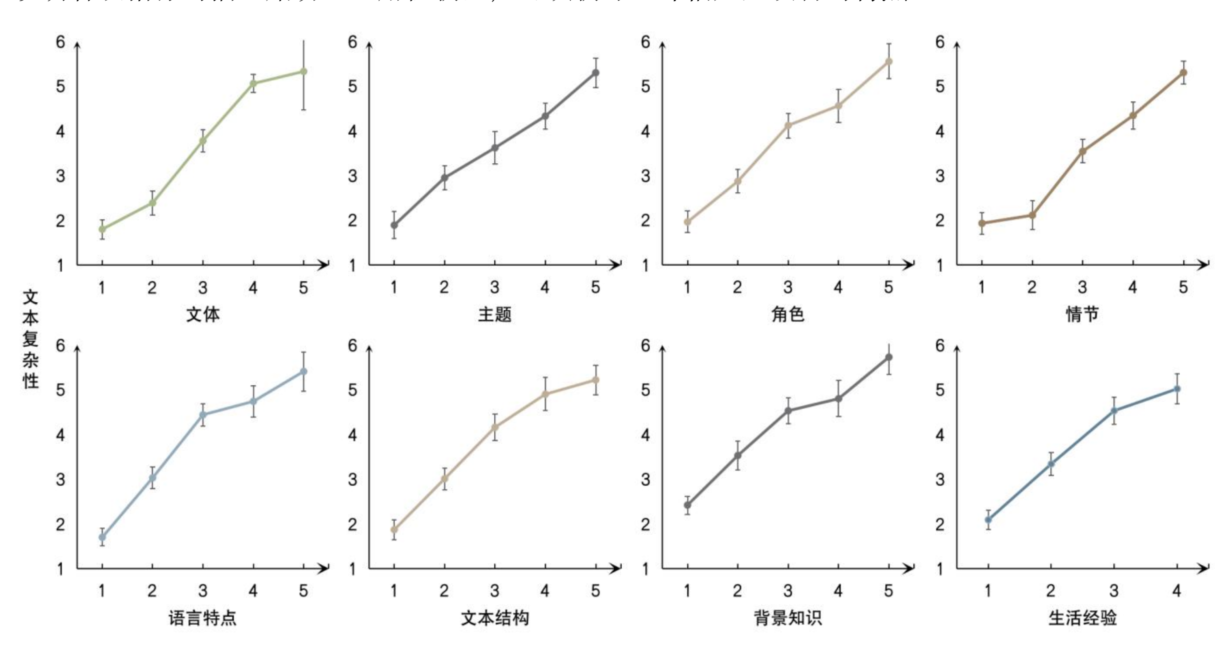
验	验，如衣食住行、儿童成长等生活化，好理解的情景。	年代差异等，或因奇幻背景产生的经历差异，但不影响整体理解。	的人生经历或社会经验，大部分内容涉及地域、文化、年代等差异，增加了理解难度。	生体验和深刻感悟，大部分内容涉及地域、文化、年代等差异，增加了整体理解的难度。		

注：评分为 1 – 5 或不适用。虽然上述指标及等级描述已被多次修订，但由于儿童读物内容丰富，等级描述或有不足、不符之处。因此，我们允许评分者在评估指标水平时填写“不适用”，但需在备注中注明不适用的原因。

（三）质性分级标准的实验验证结果

1. 评价指标的区分效果

读物分级标准可评估每本儿童读物在 8 个指标上的表现，即儿童读物的多维画像。因此，图 2 呈现了每个指标的复杂性等级分布，来作为各评价指标的区分效果。结果显示，随着指标等级的增加，文本复杂性均呈现出明显的上升趋势，说明这 8 项质性指标的等级与文本复杂性均呈单调递增关系，即简单的儿童读物的质性指标等级较低，复杂的儿童读物的质性指标等级较高，表明质性指标对文本复杂性具有较高的区分效果。另外，评分结果中的文体、语言特点、生活经验等指标被评为不适用的次数≤1 次，而不适用比例相对最高的指标为角色（仅为 2.3%）、情节（仅为 1.8%），说明各项指标的描述贴切、适用性较强，可以较好地评估儿童读物的特点。



注：误差棒为标准误的 95%置信区间，可用于直观比较指标等级在文本复杂性上的差异显著性。

图 2 质性分级指标与文本复杂性的关系

2. 评分者一致性信度

为验证质性分级标准的客观性，本研究估计了 11 组评分结果的评分者间一致性信度（表 3）。各项指标的平均皮尔逊积差相关为 0.71，平均斯皮尔曼等级相关为 0.71，均说明两位评分者参照同一分级标准评估儿童读物的各项指标时，所得结果的一致程度较高，评分结果具有较高的客观性。

chinaXiv:202301.00068v1

另外, 8 项指标的 Cronbach alpha 信度达到了 0.95, 说明各项指标结果的评分者内一致性信度也较高。

表 3 双评实验中各项指标的评分者间一致性信度

指标	皮尔逊积差相关	斯皮尔曼等级相关
文体	0.77	0.77
主题	0.65	0.62
角色	0.67	0.68
情节	0.73	0.71
语言特点	0.77	0.80
文本结构	0.72	0.72
文化知识	0.68	0.63
生活经验	0.71	0.71
平均	0.71	0.71

chinaXiv:202301.00068v1

3. 效度验证

最后，为了验证分级标准的预测效度，本研究采用多元回归法来探索每本儿童读物的平均文本指标水平对其复杂性的预测效度。结果显示多元回归模型显著, 具体经验公式为: 文本复杂性 = 0.34×文体 − 0.25×主题 + 0.17×角色 − 0.04×情节 + 0.31×语言特点 + 0.41×文本结构 + 0.26×文化知识 + 0.33×生活经验 − 0.12。此模型预测值对文本复杂性的解释率达到了 71.5%, $F(8,141) = 43.67$, $p < 0.001$, 说明指标体系对文本复杂性具有良好的预测效度。另外，我们发现当增加对一本书的评价者人数时，回归模型的解释性会有明显的提升。这表明质性评级的主观性还可以通过增加评分者人数来降低。

四、讨论

汉语儿童读物的分级具有较高的教育和应用价值。为了解决儿童阅读分级研究中量化文本指标浅表、忽视儿童认知发展需求等不足之处，本研究综合国内外相关研究，采用元分析归纳和筛选出由浅至深的多层次质性指标。在此基础上，研究在国际上首次采用 SOLO 分类评价法来编制符合儿童认知发展结构的等级水平描述，编制了《汉语儿童读物的质性分级标准》及操作手册。进一步实验结果显示，《分级标准》及文本指标具有较好的区分效果、评分者一致性信度和预测效度，能够作为分级阅读教育和出版的有效工具和科学保障。简言之，本分级标准较好地综合了影响阅读复杂性的深层次指标，可用于形成大规模分级读物库，助力教师、家长和出版从业者推广分级阅读教育，以科学推动、整体提高儿童阅读能力。

通过综合参考和整理多项国内外儿童读物质性分级研究，本研究采用元分析法归纳和分析出 8 个适用于汉语儿童读物的分级指标，即篇章层面的文体、主题、角色、情节指标，语言层面的语言特点、文本结构指标，以及背景知识需求层面的文化知识、生活经验指标。这些指标较为综合地反映了儿童读物中的语篇特征、深层思想特征和对儿童群体的认知、心理发展需求，能够客观、稳定地预测儿童读物的复杂性等级。研究结果支持了建构整合模型（construction-integration model）[25]

的观点，即阅读理解的心理表征包含建构过程（自动激活文本信息和读者的相关背景知识）和整合过程（人脑中会大量激活相关概念的联结，同时抑制无关或弱相关概念的联结）。这两种过程会反复交互来形成整体的理解，即文本思想和背景知识的有效联结。因此，本研究较好地解决了基于字词解码等自动化特征的量化研究难以真正解释儿童遇到的阅读困难，也较难给出有针对性、有价值的阅读提升建议的问题。同时，本研究筛选出的 8 个质性指标不但可作为文本复杂性的有效维度，还可作为语言学量化研究的对象[8, 26-27]和教学策略的参考，为文本复杂性向实证应用提供新的研究视角。

与前人研究不同的是，本研究在国际上首次应用 SOLO 分类法来制定出指标的等级水平描述。实验结果显示 8 个指标的等级均具有循序渐进、逐步发展的特点，说明基于 SOLO 分类法形成的等级描述确实刻画了儿童读物复杂性的普遍阶段和结构，增强了质性指标对读物复杂性的区分效果。同时，从学习角度来说，应用 SOLO 分类法的分级标准更符合儿童的语言和认知发展规律，可以解释学生在阅读任务上的外部结果和内部原因。更为重要的是，有研究指出 SOLO 分类法还能够将学生认知发展阶段和学习循环特征联系起来，诊断学生当前水平和教学目标的差距[24]。因此，SOLO 分类法或许还是在阅读教学中创设最近发展区、构建形成性测评的有效手段。

在指标筛选、等级水平描述两个环节的基础上，本研究最终编制了《汉语儿童读物的质性分级标准》及操作手册。考虑到以往的质性分级指标体系存在缺少有效性证据的问题，本研究实施了严谨的双评实验，发现良好的评分者间一致性信度、评分者内一致性信度和预测效度，说明本分级标准能够提供客观可靠、科学有效的分级结果，具有较高的可操作性。这些发现可以推动教师、出版从业人员等使用者在出版和教学实践中广泛应用分级阅读。例如，阅读分级标准可作为推广分级阅读的科学参考和教学工具，让教师参考本分级标准来为学生制定阅读书单、形成分级书库；让儿童读物出版行业规范儿童读物的制作和评价流程，产出符合儿童认知发展和阅读需求的分级读物。

当前的研究还存在许多未尽之处。比如，儿童读物的可读性和复杂性是文本相对独立的两个属性，因而一个更完善的汉语儿童读物分级体系应该是质性和量化指标及分析的有机结合。因此，下一步工作应当研究这两类指标的关系，并将其融为一体，形成更为全面、精准的分级指标体系。第二，相较于量化分级的高信度、低成本，质性分级中难免存在评分员主观性、耗时耗力等问题。但随着 GPT – 3[28]和清源中文预训练模型 CPM[29]等自回归语言模型的提出，神经网络模型的文本理解、知识概括、文本生成能力得到了进一步提升，对汉语儿童读物中深层次认知指标的自动提取、大规模儿童读物自动分级都已成为可能。因此，如何利用超大规模的 NLP 模型来完成汉语儿童读物的自动分级将是下一步研究的重要方向。

参考文献

[1] 袁曦临，戴琦. 中文儿童读本阅读接受度模型构建及其测评研究[J/OL]. 图书馆建设: 1-12.

[2] AMENDUM S J, CONRADI K, HIEBERT E. Does text complexity matter in the elementary grades? A research synthesis of text difficulty and elementary students' reading fluency and comprehension. Educational Psychology Review, 2018, 30(1), 121-151.

[3] 王蕾，毛莉. 儿童阅读素养框架体系建构研究:以统编小学语文教科书为例[J]. 课程.教材.教法,

chinaXiv:202301.00068v1

2021,7(41): 99-105.
[4] 教育部基础教育课程教材发展中心. 中小学生阅读指导目录(2020 年版)[S]. 北京: 教育部基础教育课程发展中心, 2020.
[5] 儿童青少年分级阅读内容选择标准[J].人民教育,2009(Z2):78-79.
[6] 徐默凡. 文本难度分级标准的研制和说明[J]. 上海课程教学研究, 2020(1): 49-56.
[7] CROSSLEY S A, SKALICKY S, DASCALU M. Moving beyond classic readability formulas: new methods and new models[J/OL]. Journal of Research in Reading, 2019, 42(3-4): 541-561. DOI:10.1111/1467-9817.12283.
[8] 刘苗苗，李燕，王欣萌，等. 分级阅读初探:基于小学教材的汉语可读性公式研究[J]. 语言文字应用, 2021(2): 116-126.
[9] BEGENCY J C, GREENE D J. Can readability formulas be used to successfully gauge difficulty of reading materials? Psychology in the Schools, 2009, 51(2), 198-215.
[10] MESMER H A, CUNNINGHAM J W, HIEBERT E H. Toward a Theoretical Model of Text Complexity for the Early Grades: Learning From the Past, Anticipating the Future[J]. Reading Research Quarterly, 2012, 47(3): 235-258.
[11] ELLEMAN A M, OSLUND E L. Reading Comprehension Research: Implications for Practice and Policy[J]. Policy Insights from the Behavioral and Brain Sciences, 2019, 6(1): 3-11.
[12]SHU H, CHEN X, ANDERSON R C, et al. Properties of School Chinese: Implications for Learning to Read[J]. Child Development, 2003, 74(1): 27-47.
[13] 李虹，刘苗苗，王欣萌，等. 统编小学语文教科书识字表效用分析[J]. 课程.教材.教法, 2021, 41(5): 61-66.
[14] NATION P. Teaching and learning vocabulary. In Handbook of Practical Second Language Teaching and Learning (pp. 397-408). Routledge, 1990.
[15] LAUFER B. The lexical plight in second language reading: Words you don't know, words you think you know and words you can't guess. In J. Coady & T. Huckin (Eds.), Second language vocabulary acquisition: A rationale for pedagogy (pp. 20–34). Cambridge, England: Cambridge University Press, 1997.
[16] PEARSON P D, HIEBERT E H. The State of the Field: Qualitative Analyses of Text Complexity[J]. The Elementary School Journal, 2014, 115(2): 161-183.
[17] COMMON CORE STATE STANDARDS INITIATIVE. Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects: Appendix A[S]. Washington, DC: 2010.
[18] HESS K, BIGGAM S. A discussion of "increasing text complexity". Published by the New Hampshire, Rhode Island, and Vermont departments of education as part of the New England Common Assessment Program (NECAP)[EB/OL]. (2004)[2022-12-19]. http://www.nciea.org/publications/TextComplexity_KH05.pdf
[19] American College Testing (ACT), Inc. Reading between the Lines: What the ACT Reveals about College Readiness in Reading[M/OL]// ACT, Inc, 2006[2022-10-05].
[20] 王蔷，陈则航. 中国中小学生英语分级阅读标准(实验稿)[M]. 外语教学与研究出版社, 2016.
[21] 罗德红，余婧. 儿童分级阅读研究的中美对比分析[J]. 图书馆, 2013(2): 34-37.
[22] 张沿沿，冯友梅，顾建军，等. 从知识结构与思维结构看思维评价：基于皮亚杰发生认识论知识观的演绎[J]. 电化教育研究, 2020, 41(06): 33-38.
[23] BIGGS J B, COLLIS K F. Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome)[M]. Academic Press, 2014.
[24] 蔡永红. SOLO 分类理论及其在教学中的应用[J]. 教师教育研究, 2006(1): 34-40.
[25] KINTSCH W. The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model[M]//STELMACH G E, VROON P A. Advances in Psychology: V 79. North-Holland, 1991: 107-153.
[26] 柏晓鹏，吉伶俐. 篇章结构特征对文本可读性的影响[J]. 语言文字应用, 2022(3): 62-72.
[27] TOYAMA Y. What Makes Reading Difficult? An Investigation of the Contributions of Passage, Task, and Reader Characteristics on Comprehension Performance[J]. Reading Research Quarterly, 2021, 56(4): 633-642.
[28] FLORIDI L, CHIRIATTI M. GPT-3: Its Nature, Scope, Limits, and Consequences[J/OL]. Minds and Machines, 2020, 30(4): 681-694.
[29] ZHANG Z, HAN X, ZHOU H, et al. CPM: A large-scale generative Chinese Pre-trained language model[J/OL]. AI Open, 2021, 2: 93-99.

chinaXiv:202301.00068v1

chinaXiv:202301.00068v1

Qualitative Grading Standard for Chinese Children's Books

Abstract: Elementary school students are at a critical stage of learning to read, and their language ability and cognitive development require a large amount of reading materials that are appropriately difficult and conducive to comprehension. Current leveled reading mainly relies on shallow and quantitative text indicators, such as characters, words, and sentence level lexical properties, which is less suitable for the characteristics of Chinese ideographs and children's cognitive development needs. This study first used meta-analysis to glean eight qualitative cognitive indicators, including genre, theme, character, storyline, language feature, text structure, background knowledge, and life experience. Then we used the SOLO taxonomy to develop a standardized grading description for each indicator that satisfied the developmental stages of children's cognitive ability. Finally, we constructed the *Qualitative Grading Standard for Chinese Children's Books* and its operation manual. The follow-up empirical study showed that the qualitative indicators had high discriminability, raters' consistency reliability and predictive validity, and thus could be used to evaluate children's books in an objective, reliable and valid way. In short, our Standard will help school teachers, book editors, and parents determine the appropriate grade level for children's books and use Chinese leveled reading to promote children's language and cognitive development so that they can achieve independent reading as early as possible.

Keywords: leveled reading; children's books; text complexity; qualitative indicators; SOLO taxonomy